Cell, Vol. S116, S89–S92, January 23, 2004 Copyright ©2004 by Cell Press

A Short History of a Short RNA

Commentary

Rosalind Lee,[1] Rhonda Feinbaum,[2] and Victor Ambros[1,*]
[1]Department of Genetics
Dartmouth Medical School
Hanover, New Hampshire 03755
[2]Department of Genetics
Harvard Medical School
Department of Molecular Biology
Massachusetts General Hospital
Boston, Massachusetts 02114

The intellectual backdrop motivating our effort to clone *lin-4* (Lee et al., 1993) had nothing to do with questions about noncoding RNAs or antisense gene regulation. We were simply curious about an interesting worm mutant, and everything we found out about it was unexpected. We consider ourselves very lucky to happen to have chosen *lin-4* to study. In fact, good fortune appeared at many steps before and during our *lin-4* project, often through the contributions of other people.

lin-4 was discovered in Sydney Brenner's lab in the mid 1970s through the isolation of a mutation (*e912*). The remarkable developmental defects of *lin-4(e912)* were first described by Horvitz and Sulston (1980) and characterized in detail by Chalfie et al. (1981). *lin-4(e912)* animals look terrible: they grow into long, thin "adults" with a larval skin, and they fail to stop molting at the normal stage and thus undergo extra larval stages. Chalfie et al. (1981) showed that *e912* hermaphrodites and males are completely missing many of the cell types and morphological structures typical of the wild-type adults, and instead contain many extra copies of cells ordinarily produced only at an early larval stage. It appeared that the *e912* mutation was causing a failure of temporal developmental switches throughout the animal, indicating that *lin-4* might encode a master regulator of developmental timing.

For us, a particularly alluring feature of *lin-4* was its genetic relationship with *lin-14*. *lin-14* was discovered by Edwin (Chip) Ferguson, a graduate student in Bob Horvitz's lab. Chip was characterizing genetic pathways controlling steps in development of the *C. elegans* hermaphrodite vulva (Ferguson et al., 1987). *lin-4(e912)* hermaphrodites lack even a hint of a vulva (owing to their failure to generate appropriately specified vulva precursor cells) and hence are unable to lay their eggs (which consequently hatch inside their mother and consume her). While growing cultures of *lin-4(e912)* animals for genetic experiments, Chip serendipitously discovered a spontaneous revertant that displayed nearly normal morphology and egg-laying. Chip determined that the responsible suppressor mutation was in a previously unknown gene, *lin-14*. Later, Victor, as a postdoc working in the Horvitz lab, identified apparent null alleles of *lin-14* and found that these all had developmental timing defects opposite to those of *lin-4(e912)*. Precisely the same cell lineages that reiterated early programs at later larval stages in *lin-4(e912)* animals instead completely deleted their entire early larval programs in animals lacking *lin-14* (Ambros and Horvitz, 1984).

These opposite developmental timing defects of *lin-4* and *lin-14* mutants, and the fact that loss of *lin-14* is epistatic to *lin-4(e912)*, suggested that the *lin-4(e912)* mutation resulted in an excess of *lin-14* activity. So *lin-4* might encode a *trans*-acting negative regulator of *lin-14*. This view was reinforced by another oddly lucky event: while screening worm populations for an entirely different class of mutant, Victor was startled to find an animal that looked exactly like the very distinctive *lin-4(e912)* animals! This fortuitous new mutation, *n355*, was a dominant allele of *lin-14*. Constitutive activity of *lin-14* resulted in the same collection of retarded developmental timing defects as loss of *lin-4*. A quick look through the Horvitz lab collection of egg-laying mutants isolated by Nancy Tsung revealed a second *lin-14* gain-of-function allele, *n536*.

Victor worked with Gary Ruvkun to clone *lin-14* in the Horvitz lab (Ruvkun et al., 1989), and Gary went on to sequence the *lin-14* (a novel nuclear protein) gene in his own lab at MGH (Ruvkun and Giusto, 1989). Gary's lab discovered that the *n355* and *n536* gain-of-function mutations are deletions in the 3′ untranslated region (UTR) of the *lin-14* mRNA, and that LIN-14 protein level is posttranscriptionally downregulated during worm development (Wightman et al., 1991). Therefore, if *lin-4* were involved in the temporal regulation of *lin-14*, it would probably do so via the *lin-14* 3′ UTR.

Despite the intriguing correspondence between *lin-14* and *lin-4* mutant phenotypes, we were not really sure that the cloning and molecular characterization of *lin-4* would be a worthwhile project, because the *lin-4(e912)* mutation was the only known mutant allele of *lin-4*. If *lin-4* were a normal worm gene, we knew that knockout alleles should have been more easily recovered in screens for egg-laying defective mutants. Perhaps *e912* was not a simple loss-of-function mutation in a regulatory gene. What if *e912* were a rare, arcane genetic rearrangement that disturbed development in a fashion unrelated to normal gene activity? In that case, molecular analysis of e912 would not teach us anything fundamental about normal development. On the other hand, an optimistic view was that *lin-4* might be a very small, but otherwise conventional, gene. A gene encoding a very short protein might present a very small target for mutagenesis, explaining the scarcity of *lin-4* loss-of-function alleles. We do not recall thinking that *lin-4* might encode a small regulatory RNA until much later on, when our data finally forced us to consider the possibility.

The *lin-4* cloning project in Victor's Harvard lab began in the summer of 1988, when a postdoctoral fellow, Xianjie Yang, conducted genetic mapping experiments with chromosomes polymorphic for RFLPs in the *lin-4* region. Xianjie shifted to other pursuits at the end of 1988, and Rosalind (Candy) took over the *lin-4* cloning

*Correspondence: victor.ambros@dartmouth.edu

project in January, 1989. Using the recombinant chromosomes that Xianjie had generated, Candy mapped the physical location of *lin-4(e912)* to a several hundred kilobase genomic interval. Candy then began to probe Southern blots of DNA from *lin-4(e912)* animals to determine if a DNA lesion could be identified associated with the mutation. To assay large regions of genomic DNA, she purified yeast artificial chromosome clones and used entire YACs as probes to Southern blots. These blots often produced a blizzard of hybridizing bands, but amazingly, Candy found a lesion in *e912* DNA as a missing 5 kb EcoRI restriction fragment detected in wild-type DNA by one particular YAC probe. Candy then used that YAC as the basis for identifying smaller clones that also detected the 5 kb band altered by *e912*.

In the fall of 1989, Rhonda joined the project and initiated a genetic analysis of *lin-4*, including screens for new alleles, which we felt would be critical for ultimately pinning down the gene. Although the long shot gamble of attempting to find a gene using only a single allele seemed to be paying off, there was a problem: Candy found that the *e912* lesion was complicated and involved both the deletion and rearrangement of genomic sequences. A large lesion such as this could have affected multiple genes, and so we were faced with the possibility that the *lin-4(e912)* phenotype could have been caused by the combination of mutations in multiple genes. Rhonda set about to address this issue by testing for transformation rescue of the *lin-4(e912)* phenotype using the set of smaller genomic DNA clones that Candy had isolated. Rhonda found that one of these clones, corresponding to a 3.2 kb restriction fragment (2DCla), rescued *lin-4(e912)* completely. Although the *e912* lesion extended well beyond the boundaries of 2DCla, these 3.2 kilobases had to contain *lin-4*. Rhonda made several Bal-31 deletions to pinpoint the rescuing fragment more precisely.

These were the days before complete annotated genomic sequences or automated sequencing, so Candy sequenced 2DCla by primer walking and probed cDNA libraries with 2DCla to identify candidate *lin-4* open reading frames. Candy found a cDNA overlapping 2DCla, and so for a brief time, the predicted open reading frame of that cDNA was elevated to the status of a putative *lin-4* protein. But little of that cDNA was contained in 2DCla, making it a doubtful source of 2DCla's rescuing activity. Moreover, another clone that overlapped 2DCla by only a few hundred base pairs also rescued *lin-4*. We were forced to admit that *lin-4* probably was contained in that very small region. When we found that a 700 bp Sal fragment rescued *lin-4(e912)*, we began to think that *lin-4* was odd indeed: the 700 bp Sal fragment contained all of *lin-4*, and yet no respectable ORFs were predicted (no matter how many times we resequenced it!).

At this point we began to suspect that the *lin-4* gene product might be a noncoding RNA, but admittedly, the 700 bp rescuing sequence did contain some very short open reading frames (although these lacked initiator codons and/or proper stop signals). To unequivocally determine whether any of these putative tiny polypeptides mattered for *lin-4* function, Candy cloned and sequenced functional *lin-4* from the genomes of three other *Caenorhabditis* species and found that all but a few of the putative *lin-4* micropepetides were eliminated in one or more of these other species. For the remaining putative ORFs, Rhonda introduced frameshift mutations into the *C. elegans* sequence and found that *lin-4* function was unperturbed. So, once and for all, we were sure that *lin-4* could not encode a protein. In late 1991, we began to assay for a *lin-4* transcript by RNase protection, using probes from both strands of the whole 700 bp *lin-4* sequence. Rhonda had to take a break for maternity leave in January, 1992, so Victor took over the RNase protection experiments.

We were definitely *not* expecting to find a transcript as short as 22 nt, and so we missed it for quite a while. We saw a protected species (first in February of 1992) about 60 nt in length (lin-4L), and even that seemed incredibly short for a real gene product. By mid-March of 1992, we knew the approximate position of sequences transcribed into lin-4L, and that lin-4L was predicted to form a hairpin. We began to think that the 60 nt hairpin could be the negative regulator of LIN-14 gene expression. Victor reported the RNase protection results at an informal "tea-associated research talk (TART)" with Margaret Baron's lab, and Margaret suggested that we should take more seriously the whopping protected signal at the bottom of the gel, running at around 20 nt. Despite its relative abundance compared to lin-4L, we had been inclined to dismiss this very small material as probe-specific background (although we had to admit that it was absent in samples from *lin-4(e912)*!). Margaret's comment prompted us to consider that perhaps the small material represented a real *lin-4* transcript after all. So, when our RNase protection experiments finally confirmed *lin-4S* clearly in May of 1992, we realized that the major *lin-4* gene product was ridiculously small—about 20 nt.

We were still troubled by the existence of only one known *lin-4* mutant allele, the complex *e912* aberration. If *lin-4* were a single gene residing in the 700 bp Sal region, then it ought to be possible to find a point mutation in that sequence that would cause a phenotype like that of *e912*. Here's where Rhonda's screen for new EMS-induced alleles of *lin-4* paid off: she had identified *lin-4(ma161)* by its failure to complement *e912*, and sequencing showed that *ma161* is a single base pair change within the lin-4S sequence. This reinforced our conviction that *lin-4* was a single gene and that lin-4S was almost certainly its functional product.

How and when was the complementarity to *lin-14* noticed? In our minds, an antisense RNA hypothesis grew from our proof that *lin-4* could not encode a protein, and this was supported by reading a report in *Cell* (Hildebrandt and Nellen, 1992) about a case of natural antisense in a eukaryote, *Dictyostelium*. Most importantly, however, Gary Ruvkun's lab had identified evolutionally conserved sequences in the 3′ UTR of *lin-14* in a region of the mRNA responsible for the downregulation of LIN-14; we and Gary's lab knew that these sequences could contain the elements through which *lin-4* acts. Gary shared his *lin-14* UTR sequences with us, and we sent the *lin-4* sequences to Gary. On precisely the same day in June of 1992, Victor and Gary independently noticed the antisense complementarity between *lin-4* and *lin-14*. Victor immediately called Gary; each of them read the complementary sequences to the other over the

phone, practically in unison. That was a very happy shared moment.

Victor reported the *lin-4* noncoding RNA and its complementarity to *lin-14* at the Molecular Genetics Gordon Conference in the summer of 1992, and at that meeting, Ben Lewin approached Victor: "When the story's ready, send it to us." But a major obstacle to completing the work and writing it up was that Victor and Candy had to move the lab from Harvard to Dartmouth in the summer of 1992, and Rhonda was unable to move for family reasons. Gary Ruvkun kindly provided Rhonda space in his lab at MGH, and Rhonda spent the winter of 1992/1993 there mapping more precisely the ends of the *lin-4* transcripts by nuclease protection and primer extension. Rhonda also developed a Northern blot assay for lin-4S, critical for confirming that the transcripts were not significantly modified at their ends; Victor was haunted by the possibility that *lin-4* RNA could, perversely, be covalently linked to a protein, like the poliovirus RNA of his PhD thesis (Pettersson et al., 1978). Candy pressed forward with various hybridization strategies for cloning *lin-4* sequences from other animals, including other nematodes. These attempts were unsuccessful, because, although we now know that even mammals have *lin-4*-related miRNAs (Lagos-Quintana et al., 2003), they are too divergent in sequence to have been identified by hybridization with *lin-4* probe.

We began writing in early 1993, and Gary and Victor submitted our respective manuscripts to *Cell* on the same day in early August. After the inevitable back-and-forth with the editors, the manuscripts were accepted (despite our manuscript being riddled with a particularly annoying grammatical error, prompting Ben Lewin to write, "I am reaching a point of irritation with 'it's' with the inappropriate apostrophe that may lead to the rejection of papers just on the grounds of grammatical insufficiency!").

Then came real trouble: Rhonda and Candy had decided that it was essential that their collaboration be recognized by an "equal contribution" notice on the title page of the paper. Victor had naively assumed that this would be routine, but after the paper was accepted, the *Cell* editors notified us that it was *Cell*'s long-standing policy that "equal contribution" notices must be placed in the Acknowledgments, and never on the title page. We sent *Cell* what we thought was a beautifully persuasive letter urging a change of policy. Apparently, Ben Lewin was out of the office, so in a follow-up phone call, an Associate Editor took it upon himself to cut off negotiations with the warning, "If you persist in this matter, your name will go on the list of people whose manuscripts are not welcome at *Cell*." Although we thought that this was probably a bad joke (it seemed very unlikely that *Cell* would keep such a list), we were nevertheless sufficiently intimidated to wonder what to do next. We even considered withdrawing the paper from *Cell*, although we felt obligated to stick to the plan to publish together with Wightman et al. (1993). As if by magic, with no further appeals from us, Ben Lewin wrote a few days later to tell us that *Cell* had changed its policy, and henceforth authors' equal contribution notices could be placed on the title page! This was a relief, and also a source of pride for us, to think that ours would be a seminal paper—at least with regard to this small bit of *Cell* editorial policy.

While paging through our notebooks to prepare this piece, we were astonished at how much science has changed in just ten years. All our sequencing was done by hand using 18-inch gels and autoradiography. The worm genome was still a collection of loosely arranged contigs of YACS and cosmids whose ends were unclear. The best software available for sequence alignment was GCG, which we accessed by obtuse line commands to a lethargic central mainframe. Creating the alignments in Figure 3 of our paper by hand seemed to take months—a task that would be trivial with modern software. However, by far, the most dramatic difference between now and 10 years ago for us is that in 1993 there was no interest in *lin-4* or its little RNA product outside of a very small circle of friends. Competition was nonexistent, permitting us to take the time to really do a thorough job on the *lin-4* story. We felt that an odd gene such as *lin-4* required an extra level of careful experimental proof, and it is that thoroughness and accuracy of which we are most proud. We were free to spend almost four years on the project, which is unthinkable nowadays. Today, the competitive atmosphere surrounding miRNA research forces us to publish quickly, more incrementally, and sometimes without the extra layers of proof that we would prefer.

Our *lin-4* project benefited enormously from the resources of a strong *C. elegans* research community and its sense of open communication. For example, the mutually reinforcing quality of our paper and the Wightman et al. (1993) paper is the consequence of an open sharing of unpublished data and ideas. The project would have taken far longer without the *C. elegans* physical map and clone resources (Coulson et al., 1988). The *lin-4* story is one of persistent curiosity, luck, timing, and the generosity of colleagues.

References

Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. Science *226*, 409–416.

Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of *C. elegans*. Cell *24*, 59–69.

Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. Nature *335*, 184–186.

Ferguson, E.L., Sternberg, P.W., and Horvitz, H.R. (1987). A genetic pathway for the specification of the vulval cell lineages of *Caenorhabditis elegans*. Nature *326*, 259–267.

Hildebrandt, M., and Nellen, W. (1992). Differential antisense transcription from the *Dictyostelium* EB4 gene locus: implications on antisense-mediated regulation of mRNA stability. Cell *69*, 197–204.

Horvitz, H.R., and Sulston, J.E. (1980). Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. Genetics *96*, 435–454.

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA *9*, 175–179.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843–854.

Pettersson, R.F., Ambros, V., and Baltimore, D. (1978). Identification of a protein linked to nascent poliovirus RNA and to the polyuridylic acid of negative-strand RNA. J. Virol. *27*, 357–365.

Ruvkun, G., and Giusto, J. (1989). The *Caenorhabditis elegans* heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch. Nature *338*, 313–319.

Ruvkun, G., Ambros, V., Coulson, A., Waterston, R., Sulston, J., and Horvitz, H.R. (1989). Molecular genetics of the *Caenorhabditis elegans* heterochronic gene lin-14. Genetics *121*, 501–516.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Genes Dev. *5*, 1813–1824.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell *75*, 855–862.